

Técnicas econométricas para avaliação
de impacto

**O uso de algoritmos de
emparelhamento baseados em escore
de propensão – *propensity score
matching***

Bruno César Araújo

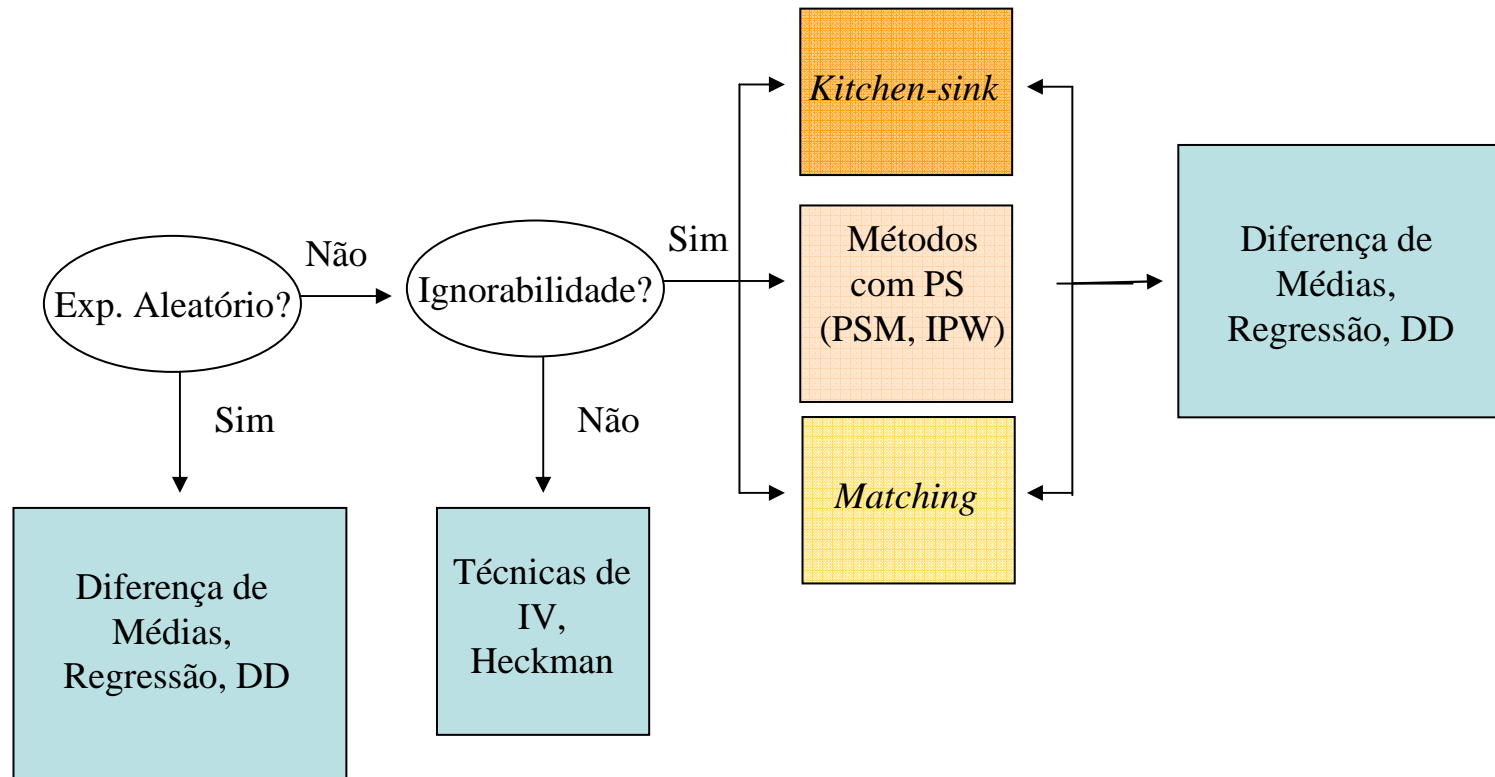
Instituto de Pesquisa Econômica Aplicada – IPEA

Brasília, 14 de maio de 2008

Pergunta da aula de hoje

Vimos na aula passada que algoritmos de *matching* podem melhorar a imputação contrafactual, contudo, as estimativas podem sofrer de viés assintótico. É possível melhorar estas estimativas, mantendo-se a hipótese de seleção em observáveis?

Esquemáticamente...



O que é e como estimar um *propensity score* ?

- Seja $D_{it}(0,1)$ um indicador se um indivíduo i recebeu um tratamento ou não no período t .
- $p_{it} (D=1)$ é a probabilidade de indivíduo i recebeu um tratamento ou não no período t , ou seja, seu *propensity score*.
- Podemos estimar $p_{it} (D=1)$ a partir de um OLS sobre observáveis, do tipo:

$$D_{it} = X_{it-1}\beta + \varepsilon_{it} \quad ?$$

O que é e como estimar um *propensity score* ?

- Em geral, não, porque:
 - 1) A probabilidade estimada não estará necessariamente entre 0 e 1, o que é contra-intuitivo;
 - 2) O modelo de probabilidade linear é essencialmente heteroscedástico (o erro de previsão do modelo é maior quando a previsão se situa em torno de 0,5).
- Entretanto, há uma classe de modelos que contornam estimando não um OLS diretamente, mas do tipo:

$$D_{it} = \Phi (X_{it-1}\beta) + \varepsilon_i,$$

Onde $\Phi (.)$ é uma função de distribuição acumulada (geralmente, a Normal – *probit* ou a logística – *logit*)

Importante: é interessante utilizar os observáveis antes do tratamento, para evitar endogeneidade com o mesmo.

Peculiaridades das estimativas *probit* ou *logit*

- Por um lado, estas estimativas limitam a probabilidade estimada entre 0 e 1 e também contornam o problema da heteroscedasticidade.
- Por outro lado, há de se prestar atenção à interpretação dos modelos e à avaliação das estatísticas tradicionais de ajuste (R^2 , por exemplo). Ainda que a inferência sobre a significância dos β e testes de restrições sobre os parâmetros sejam feitos como usualmente, o impacto marginal de uma mudança em x sobre a probabilidade NÃO é dada por β , pois ele depende também de X_i :

$$\frac{\partial p(D=1)}{\partial x} = \frac{\partial \Phi(X\beta)}{\partial (X\beta)} \cdot \beta$$

Métodos baseados em *propensity score*

- A partir do *propensity score* estimado, pode-se partir da hipótese da “ignorabilidade” do tratamento ou seleção em observáveis de maneira muito semelhante ao que procedemos anteriormente.
- *Hipótese 1*: Condicional às probabilidade $p(D=1 | X)$, D e (y_1, y_0) são independentes:

$$E(y_1 | p(D=1 | X), D) = E(y_1 | p(D=1 | X))$$
$$e E(y_0 | p(D=1 | X), D) = E(y_0 | p(D=1 | X))$$

$$ATT(X) = E(y_1 - y_0 | p(D=1 | X), D=1)$$
$$e ATE(X) = E(y_1 - y_0 | p(D=1 | X))$$

Métodos baseados em *propensity score*

- Como anteriormente, outra hipótese importante se refere à identificação: na vizinhança de um ponto, é preciso haver indivíduos que não recebam o tratamento mesmo que tenham características muito semelhantes aos que o recebem.
- *Hipótese 2*: $c < p(D=1/X=x) < 1-c$, para algum $c > 0$.

Modelos de regressão com a incorporação de PS e interações

- Tal qual anteriormente, em que incluíamos diretamente as variáveis X em um modelo de regressão, podemos substituir as variáveis e suas interações pelo *propensity score*:

$$Y_{it} = \alpha + \alpha_1 d_t + \alpha_2 d_j + \beta d_{jt} + \gamma p_{it}(D=1 | X) + \omega [p_{it}(D=1 | X) - E(p_{it}(D=1 | X))] d_{jt} + \varepsilon_{ijt}$$

O *propensity score matching*

- Em essência, o PSM faz o mesmo que o método de *hard matching*, mas com apenas uma dimensão. Isto atenua o viés assintótico da inclusão de vários observáveis.
- A aplicação da técnica se dá em quatro estágios:
 1. Estima-se o modelo probabilístico
 2. Os indivíduos são pareadas de acordo com o *propensity score*
 3. Há a checagem das chamadas condições de balanceamento ou base, isto é, se X_i é aproximadamente igual a X_j
 4. Estima-se então ATE e suas variações sobre as variáveis em nível ou em diferenças.

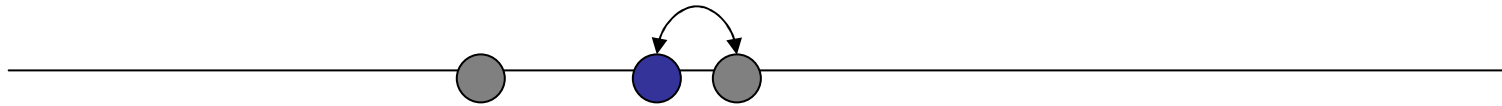
Atenção aos passos 2 e 3...

- Ao fazermos o passo 2, precisamos estar atento ao que se chama de condição de suporte comum. A idéia é que as distribuições de $p(D=1 / X)$ dos casos e controles tenham o mesmo suporte, a fim de evitar a violação da hipótese 2.
- Supõe-se que, se $p_i(.)$ é suficientemente próximo a $p_j(.)$, então X_i é aproximadamente igual a X_j . Contudo, pode acontecer que

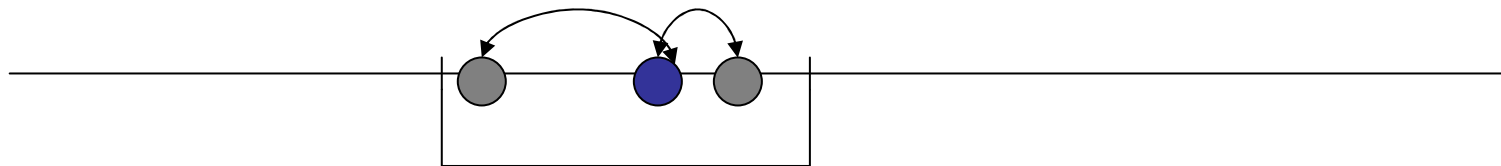
$$\Phi(X_j \beta) \approx \Phi(X_i \beta) \Rightarrow X_j \beta \approx X_i \beta \Rightarrow \sum_{k=1}^q \beta_k (x_{jk} - x_{ik}) \approx 0$$

Tipos de PSM

Nearest-neighbour

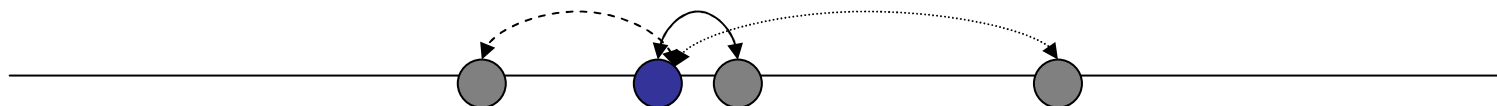


Radius matching



0,05

Kernel matching



Tipos de PSM – um balanço

- As vantagens do PSM são essencialmente as mesmas dos algoritmos de *hard matching* ou *matching* em observáveis, mais o fato de que há uma redução da dimensionalidade para o emparelhamento, o que atenua o viés assintótico no ATE.
- Contudo, embora bastante popular, o PSM não se baseia em teoria assintótica (apenas quando $p(X)$ é conhecido). Acredita-se que o PSM estime consistentemente a média das diferenças, mas pode haver problemas na estimação das variâncias.

Tipos de PSM – um balanço

- Tal qual nos algoritmos de *hard matching*, o PSM depende sobremaneira do modelo probabilístico – controvérsia Smith e Todd *vs.* Dehija e Wahba (a avaliação do ATE é especialmente sensível à especificação do modelo, ainda que as estimativas DD sejam as mais robustas neste sentido).
- Com respeito à consistência e eficiência dos estimadores, a literatura aponta que os algoritmos de PSM baseados em kernel tendem a ser mais eficientes e consistentes, pois aproveitam informação de toda a amostra.
- Há outros métodos que utilizam o *propensity score*, como o IPW (modelos *Kitchen-Sink* ponderados pelo IPW).